Characterizing the Efficacy of the NRL Network Pump in Mitigating Covert Timing Channels

Siva K. Gorantla[\$], Sachin Kadloor[\$], Negar Kiyavash[\$], Todd P. Coleman[\$], Ira S. Moskowitz[*], and Myong H. Kang[*]

Abstract

The NRL Network Pump®, or Pump, is a standard for mitigating covert channels that arise in a multi-level secure (MLS) system when a high user (HU) sends acknowledgements to a low user (LU). The issue here is that HU can encode information in the "timings" of the acknowledgements. The Pump aims at mitigating the covert timing channel by introducing buffering between HU and LU, as well as adding noise to the acknowledgment timings. We model the working of the Pump in certain situations, as a communication system with feedback and use then this perspective to derive an upper bound on the capacity of the covert channel between HU and LU in the Pump. This upper bound is presented in terms of a directed information flow over the dynamics of the system. We also present an achievable scheme that can transmit information over this channel. When the support of the noise added by Pump to acknowledgment timings is finite, the achievable rate is nonzero, i.e, infinite number of bits can be reliably communicated. If the support of the noise is infinite, the achievable rate is zero and hence finite number of bits can be communicated.

I. Introduction

A multi-level security (MLS) system stores and processes information with varying sensitivity levels in a secure and trusted manner. This requires that access to information is controlled so that no high level information can be passed to users with lower clearance levels. Thus, a crucial aspect of design of MLS systems is the prevention of communication from a user with high clearance level, HU, to a user with lower clearance level, LU, such communication in an MLS system is considered covert [1].

[\$]Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801. {sgorant2,kadloor1,colemant,kiyavash}@illinois.edu

[*]Center for High Assurance Computer Systems, Naval Research Laboratory, Washington, DC 20375. {ira.moskowitz,myong.kang}@nrl.navy.mil

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE
FEB 2012

2. REPORT TYPE

3. DATES COVERED
00-00-2012 to 00-00-2012

4. TITLE AND SUBTITLE
Characterizing the Efficacy of the NRL Network Pump in Mitigating Covert Timing Channels

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Illinois,Coordinated Science Laboratory,Urbana,IL,61801

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
IEEE Transactions on Information Forensics and Security, February 2012

14. ABSTRACT
The NRL Network Pump R , or Pump, is a standard for mitigating covert channels that arise in a multi-level secure (MLS) system when a high user (HU) sends acknowledgements to a low user (LU). The issue here is that HU can encode information in the ?timings? of the acknowledgements. The Pump aims at mitigating the covert timing channel by introducing buffering between HU and LU, as well as adding noise to the acknowledgment timings. We model the working of the Pump in certain situations, as a communication system with feedback and use then this perspective to derive an upper bound on the capacity of the covert channel between HU and LU in the Pump. This upper bound is presented in terms of a directed information flow over the dynamics of the system. We also present an achievable scheme that can transmit information over this channel. When the support of the noise added by Pump to acknowledgment timings is finite, the achievable rate is nonzero, i.e, infinite number of bits can be reliably communicated. If the support of the noise is infinite, the achievable rate is zero and hence finite number of bits can be communicated.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**26**	

Consider the communication between two users, as shown in Figure 1. The two users are labeled HU and LU, in reference to their respective clearance levels (for details, see [2]–[4]). In an MLS system, LU must be able to send packets reliably to HU and get a confirmation in the form of an acknowledgement. At the same time, any form of information flow from HU to LU, even acknowledgements, is undesirable as it can potentially be used to provide unauthorized data flow via a covert channel [1], [5], [6].

Ideally, there would be no communication from HU to LU. However, acknowledgments from HU to LU are necessary in many situations for pragmatic reasons of reliability, robustness, and system performance. Therefore, one would think that sanitizing the acknowledgements, in terms of the data they carry as much as possible, would suffice. However, the very *timing* [7] of the acknowledgements can open a communication channel between HU and LU. Thus, by using acknowledgements, the best one can hope for is a quasi-secure system, but it is still our goal to make the system as secure as we practically can. This is why one wishes to minimize the information flow through this HU to LU acknowledgement covert channel. This was the idea behind the NRL Network Pump® [2]–[4]). Here, we concentrate on the case of a single HU and a single LU.

Thus, our issue is that HU can encode information in the "timings" of the acknowledgements, and can communicate with LU at a non-zero rate. To avoid this, the Pump routes the packets and acknowledgements sent in either direction through an intermediate node, referred to as the Pump in Figure 1. Even with this intermediate node, it has been seen [2], [3], and we will further see, that covert timing communication is still possible, but at lower rates. The Network Pump provides protection by adding random noise to its acknowledgement timings to the LU [3].

In this work, we consider the problem from an information-theoretic perspective [8] and analyze the efficacy of the pump in terms of rate of information flow from HU to LU that remains available after deployment of the pump. More specifically, we show that

- Reliable communication from HU to LU is possible even after deployment of the pump in at least two scenarios: (i) if the support of random noise added by the Pump is finite, a non-zero rate is achievable; (ii) even if the support of random noise added by Pump is infinite, still reliable communication is possible albeit at zero rate. In the case of noise with finite support, we consider the example of truncated noisy channel which is used in practical implementation of Network Pump and characterize the effective rate of transfer.
- An upper bound on the capacity of the communication channel from HU to LU is derived in terms directed information. This upper bound gives the worst-case rate of information leakage between HU to LU regardless of the coding scheme employed.

The paper is organized as follows. Related work is briefly surveyed in Section II. Notation used throughout this paper is defined in Section III. In Sections IV and V, after mathematical abstraction of the pump, the covert

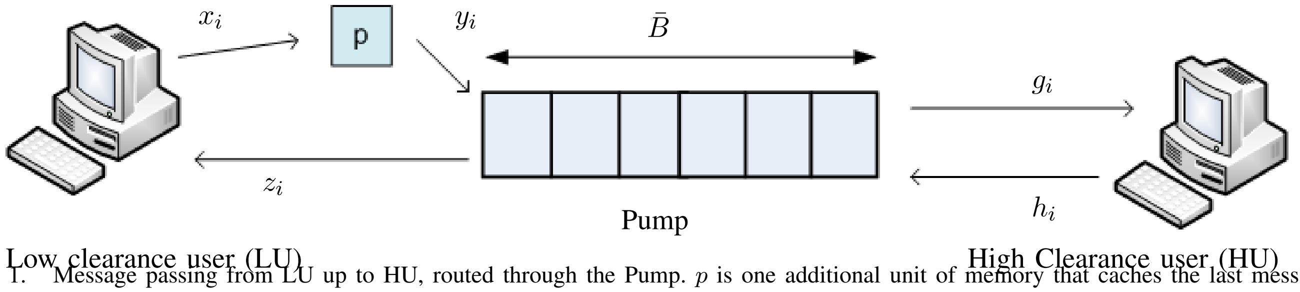


Fig. 1. Message passing from LU up to HU, routed through the Pump. p is one additional unit of memory that caches the last message sent by LU and passes onto the Pump when there is space in it.

communication over it is formulated as an instance of communication over a noisy channel with feedback problem. In Sections VI and VII, a coding scheme for communication over the pump is presented followed by an upper bound on the maximum rate of information transmission from HU to LU. Finally, we conclude in Section VIII.

II. Related Work

It is well known that having resources shared between processes can lead to covert and side channels that can leak information from one process to another. Communication channels between two processes that collude are called covert channels, one process structures its use of the shared resource in a particular way in order to communicate secret information to another. Covert channels have been studied extensively in the context of MLS systems, where they can be used to create forbidden information flows [9]–[13]. In a side channel, on the other hand, one process tries to learn something about the operation of another without the latter's cooperation.

Timing channels, channels where information is conveyed by the timings of messages, are one particular class of covert and side channels. They have been studied in the literature in variety of contexts. In cryptographic side-channels, the attacker aims at recovering cryptographic keys by utilizing the timing variations required for cryptographic operations [14], [15]. The most common mitigation technique against such channels is cryptographic blinding [15], [16]. Köpf et al. derive bounds on leakage of cryptographic operations using blinding with quantization [17].

Transmission of information by encoding it in the timings of the packets sent through a queuing system was investigated in [18]. In [19], the authors study an adversarial queuing setup, where a jammer has control of the queueing discipline. Most recently Askarov et al. adapting the periodic quantization methodology of [20] present a timing channel mitigation technique which is applicable in non stochastic arrival scenarios only. Kadloor et al. demonstrate a queuing side channel that arises in routers when first-in-first-out FIFO or Round Robin scheduling policies are used [21]. Timing channels have additionally been studied in the context of language-based security [22]–[24].

Introducing noise in timing channel is another possible mitigation approach. Perhaps the most well known example of this approach is the NRL Pump proposed for mitigating timing channels that arise in multilevel security systems (MLS) when a high confidentiality processes can communicate through Acks it sends to a low confidentiality processes [25], [26]. Past research on the Pump [2]–[4] gave rough bounds, in information theoretic terms, for the overall covert communication possible in the Pump. However, in certain situations, i.e. a *full buffer*, that analysis did not take feedback (albeit noisy feedback) into account. This paper simplifies the Pump algorithm in order to use directed information to see how feedback can influence channel capacity in the full buffer scenario. We will describe this scenario later in the paper, but we wish for the reader to keep in mind that the complete Pump algorithm incorporates such concepts as fair size to prevent the buffer from filling. However, when it comes to an MLS system, all situations must be considered before considering covert communication to be of minimal concern. This paper gives us such guidance, in certain situations, and hopefully provides further insight into the proper implementation and use of the Pump.

Besides mitigation and quantification efforts, a large volume of work on timing channels has focused on detecting timing channels [27], [28].

III. Notations and definitions

- $A^n = (A_1, A_2, \dots, A_n)$ is a vector of random variables.
- The mutual information between two sequence of random variables X^N and Y^N is defined as $I(X^N; Y^N)$. It can be expressed as (all logarithms are base 2)

$$I(X^N;Y^N) = \mathbb{E}\left[\log \frac{P_{Y^N|X^N}(Y^N|X^N)}{P_{Y^N}(Y^N)}\right]$$

- The directed information, [29], [30], is defined as

$$I(X^N \rightarrow Y^N) \overset{def}{=} \mathbb{E}\left[\sum_{i=1}^{N} \log \frac{P_{Y_i|X^i,Y^{i-1}}(Y_i|X^i,Y^{i-1})}{P_{Y_i|Y^{i-1}}(Y^i|Y^{i-1})}\right] \tag{1}$$

- For the joint random process $\{(X_i, Y_i)_{i=1}^n\}$ the directed information rate $\mathcal{I}(X \rightarrow Y)$ is defined as

$$\mathcal{I}(X \rightarrow Y) \overset{def}{=} \limsup_{n\rightarrow\infty} \frac{1}{n} I(X^n \rightarrow Y^n)$$

- $W \in \mathcal{W}$ be a message in the set of equiprobable messages transmitted from HU to LU.
- The rate R is said to be achievable using an N-length code if the message can be communicated without any error in N time units, where:

$$R \overset{def}{=} \frac{\log|\mathcal{W}|}{N} \text{ bits per transmission} \tag{2}$$

A. Directed Information

We now give a brief explanation of the role that directed information plays in communication over channels when causal feedback is available at the encoder (for a more detailed discussion see [31]). Note that for a discrete memoryless channel feedback does not increase capacity [32]. However, as we see in detail in Figure 4, the "virtual" encoder we study is physically constrained by the dynamics of the buffer, and the other physical interconnections. So the capacity of this constrained channel is unclear. Moreover, the remaining channel has memory, so feedback must be taken into account for a capacity analysis when we are in the condition of a full buffer.

Consider communication of a message W across a Shannon channel (channel inputs and outputs are governed by the distribution $P(Y^n|X^n)$) using n channel uses, without feedback. At time step i, the encoder takes the message W and transmits $X_i = e_i(W)$ across the channel. The decoder takes the channel outputs Y^n and forms an estimate of the original message $\hat{W} = d(Y^n)$. To communicate W reliably, it can be shown that the "essence" of this problem is to design $e(\cdot)$ and subsequently $d(\cdot)$ to maximize the mutual information $I(W;Y^n)$. In the absence of feedback, it can be shown that maximizing $I(W;Y^n)$ is equivalent to maximizing $I(X^n;Y^n)$.

If there is causal feedback of the outputs of the channel, then the encoder design paradigm is now

$$X_i = e_i(W, Y^{i-1}). \tag{3}$$

With feedback, $I(W;Y^n)$ can be re-written as (Figure 2):

$$\begin{aligned}
I(W;Y^n) &= \mathbb{E}\left[\log \frac{P_{Y^n|W}(Y^n|W)}{P_{Y^n}(Y^n)}\right] && (4)\\
&= \sum_{i=1}^{n} \mathbb{E}\left[\log \frac{P_{Y_i|W,Y^{i-1}}(Y_i|W,Y^{i-1})}{P_{Y_i|Y^{i-1}}(Y_i|Y^{i-1})}\right] && (5)\\
&= \sum_{i=1}^{n} \mathbb{E}\left[\log \frac{P_{Y_i|Y^{i-1},X^i,W}(Y_i|Y^{i-1},X^i,W)}{P_{Y_i|Y^{i-1}}(Y_i|Y^{i-1})}\right] && (6)\\
&= \sum_{i=1}^{n} \mathbb{E}\left[\log \frac{P_{Y_i|Y^{i-1},X^i}(Y_i|Y^{i-1},X^i)}{P_{Y_i|Y^{i-1}}(Y_i|Y^{i-1})}\right] && (7)\\
&= I(X^n \to Y^n). && (8)
\end{aligned}$$

where,

- (5) follows from the product rule.
- (6) follows because X_i is a function of W and Y^{i-1} (3).
- (7) follows because, given (X^i, Y^{i-1}), Y_i is conditionally independent of the message W.

Therefore, maximizing $I(W;Y^n)$ is equivalent to maximizing $I(X^n \to Y^n)$. Only in the case of no feedback are $I(W;Y^n)$, $I(X^n \to Y^n)$, and $I(X^n;Y^n)$ equivalent. Therefore, directed information is the function that (when maximized) characterizes the capacity of the whole channel (the original noisy channel and effect of feedback) between the original message W and output Y^n [33].

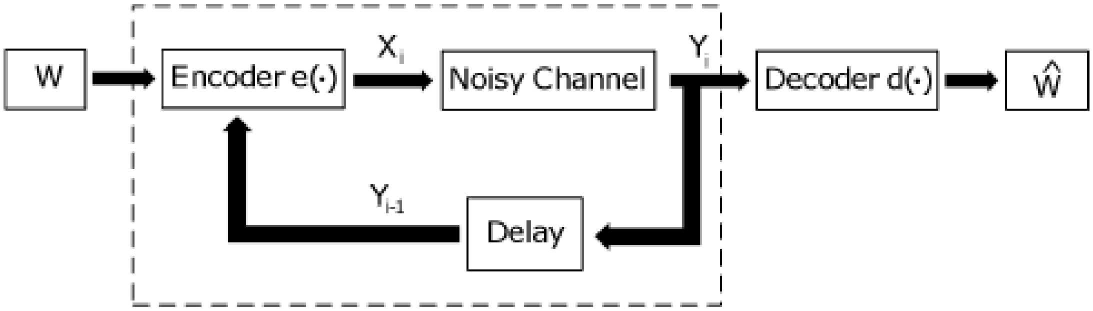


Fig. 2. A communication system over a noisy channel with feedback

IV. Working of the Pump

The buffer inside the pump is composed of two parts. There is a buffer b of size $\bar{B}$, and a buffer p of size one. Each time LU intends to send a packet to HU, it inserts a packet into the buffer p in the Pump. If the buffer b is not full, then the packet is transferred from buffer p to buffer b, and an acknowledgement (ACK-L) is sent to LU. When the buffer b is full, the packet is retained in buffer p, and no acknowledgement is sent to LU. The packet is transferred from p to b when some space clears up.

Whenever the buffer b is non-empty, the Pump forwards a packet from the buffer to HU. If HU accepts the packet, i.e., if the packet is transmitted to HU without any error, an acknowledgement (ACK-H) is sent from HU to the Pump. After receiving ACK-H, the Pump deletes the packet from the buffer.

It is important to note that the acknowledgement sent by the Pump to LU, and the acknowledgement sent by the HU to the Pump serve two different purposes. The former confirms that the buffer in the Pump is not full and that the packet has been successfully written into it, while the latter confirms a successful read by the HU. In particular, it is crucial to note that the LU **does not know** if and when the HU has read a packet. This isolation of the two users is the desired role played by the Pump.

A. A covert timing channel through the Pump

As stated in the description above, the actions of HU and the actions of LU are isolated by the Pump. However, there is one scenario in which the Pump fails to provide this isolation. This is the case when the buffer in the Pump is full. Note two things: First, the full buffer channel was studied in [3], but that was done *without feedback* being taken into consideration—so those results may be interpreted as being overly optimistic; hence the need for this paper. Secondly, the operation of the Pump in its usual mode of not having a full buffer still may leak information from HU to LU. This covert leakage has been estimated in prior Pump work, and is not the subject of this paper. Here we are only concerned with the covert channel that arises due to a full buffer.

Consider the case when LU sends a packet to the Pump when the buffer is full. In this scenario, the packet is written into buffer p. The packet will get transferred from the buffer p to buffer b only when the latter has space.

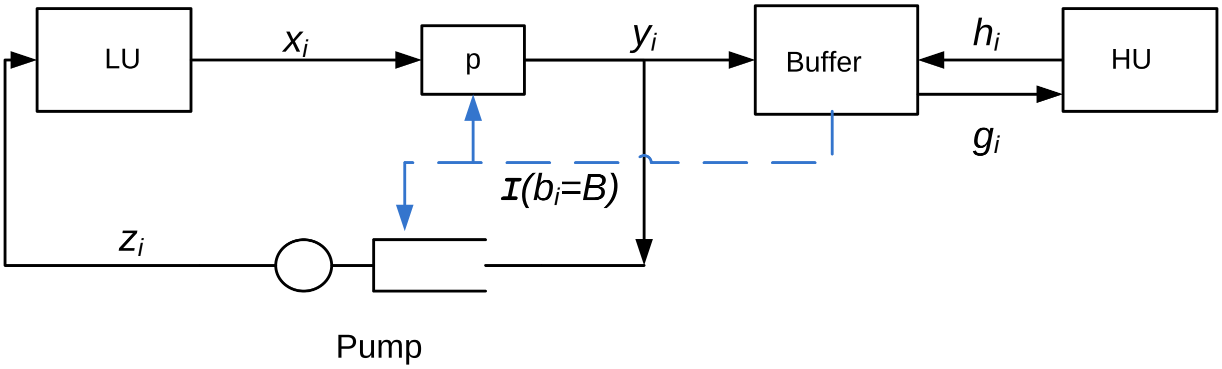


Fig. 3. A systems theory diagram denoting the joint interaction between all the observed and unobserved state variables. The arrows in black show the direction in which packets flow through this network. LU and HU are the low clearance and the high clearance user respectively. p is one additional unit of space where the last transmitted packet from LU is cached before passing it onto the Buffer inside the pump. The dashed blue arrow is drawn to indicate how the state of the buffer, whether it is full or not, influences the working of p and the noise added at the pump.

Hence, unlike when the buffer b is not full, LU does not get an acknowledgement immediately. It will receive an acknowledgement as soon as HU successfully reads a packet from the Pump, thereby deleting a packet from the buffer. Thus, LU knows exactly the timings when HU reads the packets and therefore, in this scenario, HU can communicate to LU by encoding a message in the time it sends acknowledgements to the buffer.

To avoid the above mentioned scenario, the Pump adds a random time, based on a moving average of HU acknowledgement times, to every acknowledgement that it sends to LU. We analyze a simplified version (the moving average is held constant) of this system to study if the addition of this 'noise' to the acknowledgements guarantees an acceptably minimal covert communication rate from HU to LU.

B. Dynamics of the process

The complete dynamics of the system are given below and presented in Figure 3. The thick black lines denote the direction of physical transmission of the packets and the acknowledgements. The dotted blue lines represent the virtual feedback present because of the dynamics of the system. This becomes clear from the following description of the system.

System model: We denote by b_i and p_i, the state of the respective buffers b and p at time i. The state of a buffer is the number of packets in the buffer. We assume that the time is discretized, so $i = 0, 1, 2, 3, \ldots$. We will further assume that all the links are reliable, i.e. there is no packet loss in transition.

Transmission of packets from the LU: Let x_{i-1} be the binary valued random variable indicating whether a packet was transmitted at time i. Note that we use the notation x_{i-1} and not x_i to denote the packet sent at time i as we will later interpret $\{x_i\}$ as being the feedback from LU to HU. LU sends a packet to the Pump and waits for ACK-L. It sends the next packet only after receiving ACK-L for the previous packet.

Insertion of packets into the Pump : If the LU sends a packet at time i, it is first written into the buffer p. If the buffer b is not full, then the packet is immediately written onto the buffer b and the packet is cleared from buffer p. However, if the buffer b is full, then the packet is retained in buffer p.

Acknowledgements sent to LU: If a new packet is sent by LU at time i, and if it is successfully written into the buffer b, then the Pump prepares to send ACK-L to LU. Note that a successful insertion of the packet into the buffer b happens only when the buffer is not full.

Random back-off for the transmission of acknowledgements: Once the Pump is ready to send ACK-L, it waits for a random number of time slots before it sends ACK-L out. Let us denote by q_i, the number of packets at time i which are not acknowledged by the Pump. We can think of such packets waiting in a queue, whose service times are distributed according to some distribution. Then q_i is the 'state' of that queue at time i. Once the random back-off time expires, an acknowledgement is sent to LU immediately. Let z_i denote the departure process from this queue.

Constraint on the transmissions of packets by LU: LU is not allowed to transmit the next packet to the Pump until it has received ACK-L for the packet transmitted previously. This constraint implies that q_i, which is the number of packets which are not yet acknowledged by the Pump, can at most be one.

Transmission of packets from the Pump to HU: Whenever the buffer b is non-empty, the Pump forwards a packet to HU. This is denoted by the process $\{g_i\}, g_i \in \{0,1\}$. The packet is not erased from the buffer though. The buffer waits for an acknowledgement from HU before transmitting the next packet.

Acknowledgement from HU: After the packet is successfully received by HU, it sends ACK-H to the Pump, given by $\{h_i\}, h_i \in \{0,1\}$. After receiving ACK-H, the Pump erases the packet from the buffer. It is important to note that HU controls the time when ACK-H is sent to the Pump.

V. Communication over a channel with feedback

Consider Figure 4 which captures the system model developed so far. LU sends a stream of packets $\{x_i\}$, where $x_i \in \{0,1\}$ is the indicator if a packet has been sent in time slot i.

The buffer p can be thought of as a queue, whose states are zero and one. The input to this queue happens through the arrival process $\{x_i\}$, and a departure happens whenever the buffer is not full. Denote $\{y_i\}$ as the output process of the buffer p. Thus,

$$y_i = f_y(p_{i-1}, x_{i-1}, b_i) \tag{9a}$$

$$y_i = \begin{cases} 1 & p_{i-1} = 0, x_{i-1} = 1, I(b_i = \bar{B}) = 0. \\ 1 & p_{i-1} = 1, x_{i-1} = 0, I(b_i = \bar{B}) = 0. \\ 0 & \text{otherwise.} \end{cases} \tag{9b}$$

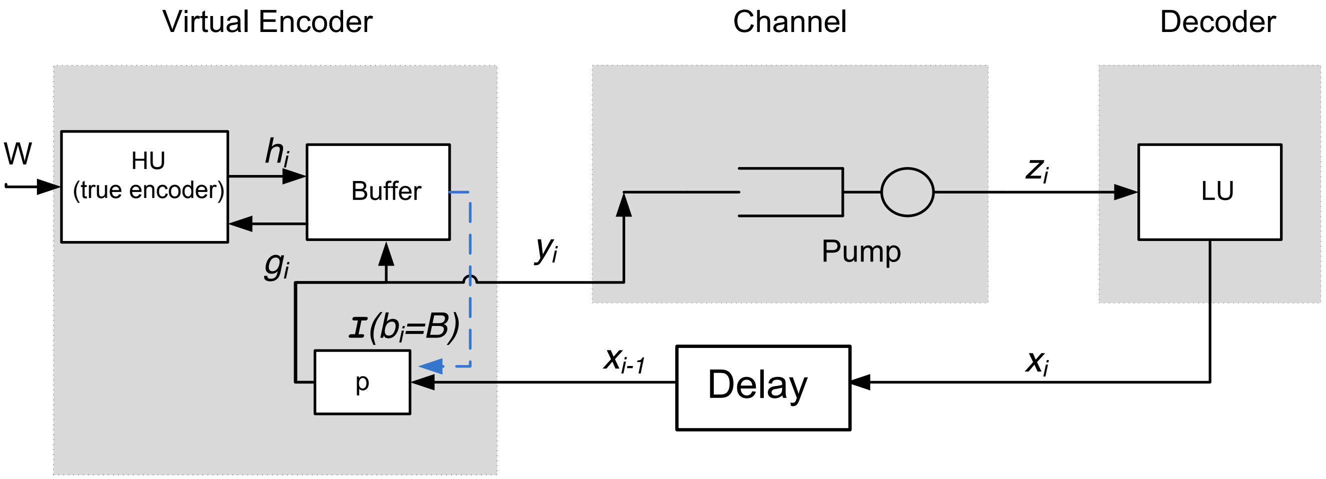


Fig. 4. System drawn as communication system over a channel with feedback as described in Sec IV-B.

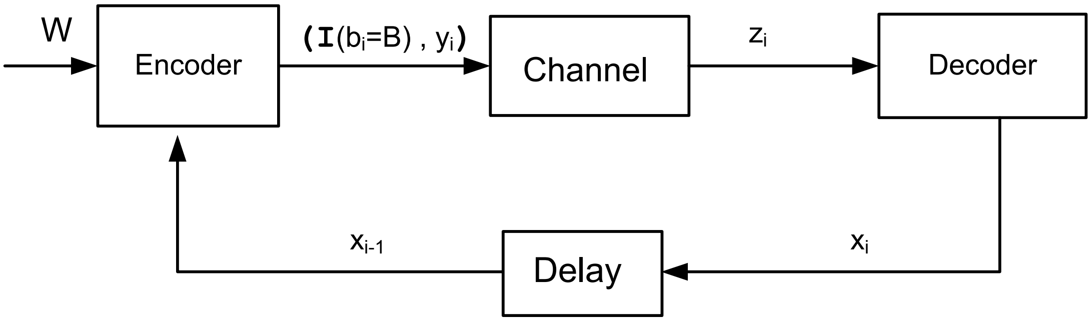


Fig. 5. A canonical communication system with feedback. The LU (Figure 4) acts as a decoder, the Pump plays the role of a noisy channel, and the dynamics of the system along with the actions of the HU act as a virtual encoder.

The number of packets in the buffer is given by b_i. The indicator function $I(b_i = \bar{B})$ is one if the buffer b is full, zero otherwise. The buffer p is a unit memory buffer. Let p_i denote the state of the buffer at time i. Dynamics at buffer p is

$$p_i = f_p(p_{i-1}, b_i, x_{i-1}) \quad (10a)$$

$$p_0 = 0 \quad (10b)$$

$$p_i = \begin{cases} 1 & p_{i-1} = 0, x_{i-1} = 1, I(b_i = \bar{B}) = 1. \\ 0 & \text{otherwise.} \end{cases} \quad (10c)$$

Note that when the buffer is not full, $y_i = x_{i-1}$. If $y_i \neq 0$, then a new packet gets written into the buffer, and b_i increments by one. The process $\{y_i\}$ serves as an input to the queue of the Pump. Inputs to buffer b are through the departure process from buffer p, and departure process through the sending of acknowledgements from HU.

Hence,

$$\begin{aligned} b_{i+1} &= f_b\left(b_i, h_i, y_i\right) && \text{(11a)} \\ &= b_i - h_i + y_i. && \text{(11b)} \end{aligned}$$

The Pump takes the process $\{y_i\}$, and after a random service time, lets out the packets, denoted by the process $\{z_i\}$. In the real system, this mimics the process of the Pump taking in data packets from LU and transmitting the acknowledgement packet after a random delay. Thus, the Pump acts as a FCFS queue with a noisy service time.

q_i represents the number of packets at time i which have not been acknowledged. The input to this queue is through the departure process from buffer p. The departure from this queue depends on whether the random back-off time has expired or not. The queue dynamics can hence be written as

$$\begin{aligned} q_{i+1} &= f_q\left(q_i, z_i, y_{i+1}\right) && \text{(12a)} \\ &= q_i - z_i + y_{i+1}, && \text{(12b)} \end{aligned}$$

where z_i is the random variable that denotes whether an acknowledgement has been sent or not.

We assume that the decoder (LU) inserts a new packet into the system as soon as it receives acknowledgements for all the packets it has transmitted so far.

$$x_i = f_x(x^{i-1}, z^i) \qquad \text{(13a)}$$

$$x_i = \begin{cases} 1 & \sum_{k=1}^{i} x_{k-1} = \sum_{k=1}^{i} z_k. \\ 0 & \text{otherwise.} \end{cases} \qquad \text{(13b)}$$

Each time the buffer is not empty, it sends a packet to HU. This is captured in the process $\{g_i\} \in \{0, 1\}$. The actions of HU is a function of the packets he receives, $\{g_i\}$, and the message W, which the HU wishes to convey to LU. It is assumed that the buffer sends a new packet to HU as soon as it has received an acknowledgement for the previous packet (and if the buffer is non-empty). The output process of the queue at HU affects number of packets in the buffer. This output process models the acknowledgement packets sent by HU, which result in packets being erased from the buffer.

$$g_i = f_g(b_i, g^{i-1}, h^{i-1}) \qquad \text{(14a)}$$

$$g_i = \begin{cases} 1 & \sum_{k=1}^{i-1} g_k = \sum_{k=1}^{i-1} h_k, \ b_i > 0. \\ 0 & \text{otherwise.} \end{cases} \qquad \text{(14b)}$$

$$h_i = f_h(g^i, W) \qquad \text{(14c)}$$

We will state a brief lemma which we will use later.

Lemma 5.1: The variables y^n and b^n are deterministic functions of z^n, given the message W and initial state of the buffer b, B_0.

Proof: Follows from equations (9),(10),(11), and (14). ■

A. Deviations from the original Pump

What we have presented above is a simplification of the actual Pump algorithm in the following ways. This has been done to examine the entire suite of possible covert communication from HU to LU, and to make sure the special cases, such as the full buffer, are adequately addressed.

The actual Pump sends messages to LU based upon an off-set and truncated exponential random variable with mean based upon a moving average of past HU acknowledgement times. In this paper, we assume that the noise added by the pump is a random variable whose statistics do not change with time. This approximation is necessary in order to quantify fundamental limits on the capacity of the covert channel. Moreover, in many communication systems, i.i.d. noise is the worst possible form of noise. We conjecture that it is the same for Pump as well, although we do not conclusively justify it.

Another deviation from the original pump setup is that this paper does not include the concept of ”Fair size” (sec. 3.1.3, [2]). The concept of **Fair Size** is a design parameter intended to keep the queue length at a desired level. This desired level will ensure that even if the flow of input packets is bursty, there is enough space for all the packets. The choice of the Fair size in [2] is subjective. This concept is not relevant in our problem formulation because in our setup, the traffic from the Low User is not bursty. Infact, the Low User is always active and assumed to have the next packet ready for transmission as soon as it receives the acknowledgement for the previous packet. Finally, we did not consider packet drops in our modeling.

B. Differences between other communication channels with feedback

Network pump is similar to trapdoor problem [34] and exponential server timing channel [35] to the extent of the following: *both the trapdoor channel and exponential server timing channel can be represented as channels with linear dynamics and internal feedback, [36], [37]. We are motivated from those examples to use directed information upper bounds for the amount of information flow through the system.* Having said that, the Network pump setup is significantly different from the above two examples in the following aspects:

In the exponential server timing channel, the queue state dynamics are linear. In the case of Network Pump, the dynamics are very complex. Compare Figure 12 of [36] to Figure 2. The input to the channel X_i in the case of exponential server timing channel is a linear function of the message W and feedback Y^{i-1}. In the case of Network Pump, the same relationship is a composition of the functions given in the equations (8),(9),(10), and (14) - which is clearly non-linear and more complex. Please note that there are differences in the notations used in [36] and this paper.

The actions of the High User and the Low User who play the role of the encoder and the decoder respectively are restricted. The High User can only acknowledge packets which have already been sent by Low User and the Low User can only send packets after it has received an acknowledgement from the pump. This is very different from the traditional communication systems in which, other than an average power/rate constraint, the possible actions of the encoder and decoder are in no way restricted by the channel.

Because of these differences, the subsequent analysis is fundamentally different.

VI. A simple transmission strategy

We present a coding scheme which illustrates that a covert communication channel can indeed be created from HU to LU. We start the analysis at time $t = 0$ when the buffer is assumed to be empty:

Phase 1: LU sends $\bar{B}$ packets addressed to HU back to back. That is, the first packet is sent from LU at $t = 0$ and an ack is received at $t = N_1$, where N_i is the noise added by the pump for the acknowledgement to the i^{th} packet. The second packet is sent at $t = N_1 + 1$ and an ack received at $t = N_1 + N_2$, and so on. The pump, after receiving the packet forwards it to the HU. However, in phase 1, HU does not ack any packet that he receives. Packets would therefore get accumulated inside the buffer of the pump. This stage lasts for $t = \bar{B}K$ time slots, where K is a coding parameter chosen later. The LU and HU assume that the buffer is full at this time. This will be true if LU has, by this time, received ACKs from the pump for all his B packets. We will shortly show that this assumption is true when K is chosen to be a suitably large value. After this time, if LU tries to send any more packets, they will get dropped.

Phase 2: At this time, HU can start communicating to LU by selecting the time when he reads packets from the buffer. LU sends one packet every time slot starting $t = \bar{B}K$. They are dropped at the pump because the buffer inside the pump is full. However, as soon as HU acks one of the $\bar{B}$ packets which are already in the buffer, the pump accepts one new packet from LU.

Starting at time $t = \bar{B}K$, HU waits for a random time M_1 before sending an ack for one of the packets already in the buffer. Hence, at time $t = \bar{B}K + M_1$, one space in the buffer clears up. Recall that LU, in Phase 2, is continuously trying to send a packet. The packet sent by LU at time $t = \bar{B}K + M_1 + 1$ is therefore written into the buffer. The pump sends an ack to LU at time $t = \bar{B}K + M_1 + 1 + N_{B+1}$.

M_1 is a discrete valued random variable. For the time being, we will assume that M_1 has finite support set. HU can choose a distribution for M_1 that will maximize the rate of information transfer. We will comment on this distribution later on.

This phase lasts for duration K as well. At time $t = (\bar{B}+1)K$, HU assumes that LU has received an ack for the $\bar{B}+1^{th}$ packet. He then waits for a random time M_2 and then reads a packet from the buffer, and the process continues as detailed in Phase 2.

In Phase 2, LU transmits packets at each time slot, HU reads packet i at time $t = (\bar{B}+i-1)K + M_i$, and LU gets an ack for the packet $\bar{B}+i$ at time $t = (\bar{B}+i-1)K + M_i + 1 + N_i$. Note that in making this statement, we have inherently assumed that $M_i + N_i < K$, we need to prove this.

1) Analysis of probability of decoding error: Consider a time horizon of $t \in [\bar{B}K+1, (\bar{B}+n)K]$ time units divided into n blocks of K time units each. Let $\{M_i \in \mathcal{M}\}_{i=1}^{n}$ take values in $\{1, 2, \cdots, |\mathcal{M}|\}$ where the support size $|\mathcal{M}|$ is finite. We consider the analysis of error probability for two different cases

- the noise N_i is geometric with mean $\mathbb{E}[N_i] = \frac{1}{\mu}$.

$$P_N(N_i = k) \quad = \quad (1-\mu)^{(k-1)}\mu, \quad k = 1, 2, 3, \cdots \tag{15}$$

- the noise N_i is truncated geometric of rate μ and truncated at K' where the support K' and duration of each block in the strategy K satisfy $K' + |\mathcal{M}| = K$.

$$Pr(N_i = k) = \begin{cases} \frac{(1-\mu)^{k-1}}{1-\mu^{N+1}}\mu & k \in \{1, 2, \ldots, K'\}. \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

If $M_i + N_i < K, \forall i \in \{1, 2, \ldots, n\}$, then the channel essentially behaves like an additive noise channel. In this scenario, the probability of decoding error can be shown to go to 0 by invoking Asymptotic Equipartition Property (AEP) [32].

Since the truncated geometric case is a special case of geometric noise, we analyze the geometric noise case first. Further analysis corresponds to geometric noise unless otherwise specified.

For the dynamics of the system given above, define

- The transmission times of the HU as $\{A_i = (\bar{B}+i-1)K + M_i\}$, and the receiver times of the LU as $D_i = \{(\bar{B}+i-1)K + M_i + 1 + N_i\}$.
- $K' = K - |\mathcal{M}|$,
- Define event E_0 to occur if the time taken in the first phase to fill the Pump to its full capacity $\bar{B}$ is less than $\bar{B}K$ time units.
- Define event $E_{1,i}$ to occur when $D_i \leq (\bar{B}+i+1)K$. Define $E_1^n = (\cap_{i=1}^{n} E_{1,i}) \cap E_0$. Event E_1^n is equivalent to saying for each i, $M_i + N_i \leq K$.
- Define event $E_{2,i}$ to occur if $N_i \leq K - |\mathcal{M}| = K'$ and $E_2^n = (\cap_{i=1}^{n} E_{2,i}) \cap E_0$. Since the support size of M_i is bounded above by $|\mathcal{M}|$, $N_i \leq K', \forall\ i$ implies $M_i + N_i \leq K, \forall\ i$. Thus event $E_{2,i}$ implies event $E_{1,i}$ and

$E_2^n \subseteq E_1$. Moreover, if event E_2^n occurs, then

$$(\bar{B}+i-1)(K) < A_i < D_i < (\bar{B}+i)K, \forall i = 1, 2, \cdots, n \tag{17}$$

and hence the communication in a given block is independent of dynamics of the past, conditioned upon event E_2^n. Hence, the proposed achievable scheme converts communication over this channel with dynamics to n-channel uses over an additive noise channel with input M_i and output $M_i + N_i$ conditioned upon event E_2^n.

Lemma 6.1: The probability of event E_1^n occurring is given by

$$\mathbb{P}\left(E_1^n\right) = (1-(1-\mu)^{K'})^n(1-(1-\mu)^{K\bar{B}}) \tag{18}$$

$$\rightarrow 1, \text{ as } n \rightarrow \infty \qquad \text{if } K' > \frac{\log(n)}{|\log(1-\mu)|} \tag{19}$$

Moreover, the conditional joint distribution of noise N_i conditioned upon event E_2^n is given by

$$\begin{aligned} &\mathbb{P}\left(N_1 = a_1, N_2 = a_2, \cdots, N_n = a_n | E_2^n\right) \\ &= \begin{cases} \frac{\prod_{i=1}^n (1-\mu)^{a_i-1}\mu}{\left((1-(1-\mu)^{K'})\right)^n}, & a_i \leq K', \forall\, i = 1, 2, \cdots, n \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{20}$$

Proof: Refer Appendix-A. ■

Lemma 6.2: For a block length n, such that each $M_i \sim P_M$, $N_i \sim P_N$ given in (15), and decoder has access to $\{M_i + N_i\}_{i=1}^n$, R given by $R \leq H(M+N) - H(N)$, there exist a block code such that

$$|\mathcal{W}| \quad \geq \quad e^{n(R-\epsilon_1(n))} \tag{21a}$$

$$\mathbb{P}\left(\hat{W} \neq W\right) \quad \leq \quad \epsilon_2(n) \tag{21b}$$

$$P_{e,1} \triangleq \mathbb{P}\left(\hat{W} \neq W | E_2^n\right) \quad \leq \quad \epsilon_3(n) \tag{21c}$$

where $\epsilon_1(n), \epsilon_2(n), \epsilon_3(n) \rightarrow 0$ as $n \rightarrow \infty$ if n and K' satisfy $K' > \frac{\log n}{|\log(1-\mu)|}$.

Proof: Refer Appendix-B. ■

Corollary 6.3: For a block length n, such that each $M_i \sim P_M$, $N_i \sim P_N$ given in (16), and decoder has access to $\{M_i + N_i\}_{i=1}^n$, R given by $R \leq H(M+N) - H(N)$, there exist a block code such that

$$|\mathcal{W}| \quad \geq \quad e^{n(R-\epsilon_1(n))} \tag{22a}$$

$$\mathbb{P}\left(\hat{W} \neq W\right) \quad \leq \quad \epsilon_2(n) \tag{22b}$$

where $\epsilon_1(n), \epsilon_2(n) \rightarrow 0$ as $n \rightarrow \infty$.

Proof: For the truncated noise given in (16), the event E_2^n occurs with probability 1, $\mathbb{P}\left(E_2^n\right) = 1$. Hence,

$$\mathbb{P}\left(\hat{W} \neq W | E_2^n\right) \;=\; \mathbb{P}\left(\hat{W} \neq W\right) \leq \epsilon_2(n) \tag{23}$$

■

With our problem setup, it is not always possible for the LU to have access to $\{M_i + N_i\}_{i=1}^n$. To avoid this we separate communication instants long enough so that LU has access to $\{M_i + N_i\}_{i=1}^n$.

Theorem 6.4: For the problem setup of Network pump, such that

- each $M_i \sim P_M$, $N_i \sim P_N$ given in (15), R given by $R \leq H(M+N) - H(N)$,
- $\tilde{n} = nK$ time-instants are used in total, where n and K satisfy

$$K \;\geq\; \frac{\log(n)}{|\log(1-\mu)|} + |\mathcal{M}| \tag{24}$$

there exist a code such that $|\mathcal{W}|$ number of messages can be distinguished where

$$|\mathcal{W}| \;\geq\; e^{n(R-\epsilon_1(n))} \tag{25a}$$

$$\mathbb{P}\left(\hat{W} \neq W\right) \;\leq\; \epsilon_4(n) \tag{25b}$$

where $\epsilon_1(n), \epsilon_4(n) \to 0$ as $n \to \infty$.

Proof: Refer Appendix-C. ■

Corollary 6.5: For the problem setup of Network pump, such that

- each $M_i \sim P_M$, $N_i \sim P_N$ given in (16), R given by $R \leq H(M+N) - H(N)$,
- $\tilde{n} = nK$ time-instants are used in total, where the support of noise K' and duration of each block in the strategy K satisfy $K' + |\mathcal{M}| = K$ as defined in (16),

there exist a code such that $|\mathcal{W}|$ number of messages can be distinguished where

$$|\mathcal{W}| \;\geq\; e^{n(R-\epsilon_1(n))} \tag{26a}$$

$$\mathbb{P}\left(\hat{W} \neq W\right) \;\leq\; \epsilon_2(n) \tag{26b}$$

where $\epsilon_1(n), \epsilon_4(n) \to 0$ as $n \to \infty$.

Proof: Since, $M_i + N_i \leq K$ for all i, event E_2 holds true with probability 1. Hence, from corollary 6.3

$$\mathbb{P}\left(\hat{W} \neq W\right) \;=\; \mathbb{P}\left(\hat{W} \neq W | E_2^n\right) = \epsilon_2(n) \tag{27}$$

■

Thus, from Thm 6.4, we can be able to distinguish between $|\mathcal{W}|$ number of messages given by (25a) with error tending to 0. The effective rate of transfer will be

$$\tilde{R} = \frac{nR}{nK} \leq \frac{R}{\frac{\log(n)}{|\log(1-\mu)|} + |\mathcal{M}|} \rightarrow 0$$

Hence, the effective rate of transfer is 0 while promising reliability. Though the support of noise is infinity, we were able to find reliable codes such that a finite number of messages can be transmitted.

For the truncated geometric noise case, where the noise support is finite, the effective rate is given by

$$\tilde{R}_{trunc} = \frac{nR}{nK} = \frac{R}{K' + |\mathcal{M}|} > 0$$

Hence, if the noise added by the pump N_i has finite support, then we can transmit infinite number of bits or equivalently a non-zero rate while promising reliability. Note that the effective rate $\tilde{R}_{trunc}$ depends on the value of $R = H(M+N) - H(N)$. While the entropy of noise N for the truncated geometric noise with parameters μ and K' is fixed, the entropy of $M+N$ can be maximized to improve the rate of transfer. We will now analyze the characteristics of this rate $\tilde{R}_{trunc}$ as a function of supports $|\mathcal{M}|$ and K'.

A. Simulation results for Truncated Noise

The mutual information maximizing input distribution for an additive noise channel with truncated geometric noise is not easy to compute in closed form. We hence resort to simulations. We use the Blahut-Arimoto algorithm (refer Chapter 10 of [32] for details) to compute it.

The rate at which information is conveyed from HU to LU is then

$$\frac{1}{K} I(M;N) = \frac{1}{|\mathcal{M}| + K'} I(M;N)$$

In Figure 6, we plot this rate as a function of the support size $\mathcal{M}$. The rate achieved is of the form $O(\frac{\log \mathcal{M}+N}{\mathcal{M}+N})$. For a given value of the support set of noise N, there is a best value of $\mathcal{M}$ which results in highest rate. If this value of $\mathcal{M}$ is used by the encoder, the rates obtained are plotted in Figure 7. We also plot the tradeoff between forward rate (from LU to HU) and the covert communication rate (from HU to LU) in Figure 8. For this plot, K' was fixed at 100, the mean of the truncated geometric variable was varied and the covert communication rate was computed. The forward rate is the inverse of the mean of the truncated geometric noise. The main focus of the paper is to demonstrate that covert communication over a network Pump is possible, and hence we do not provide a detailed analysis of this tradeoff.

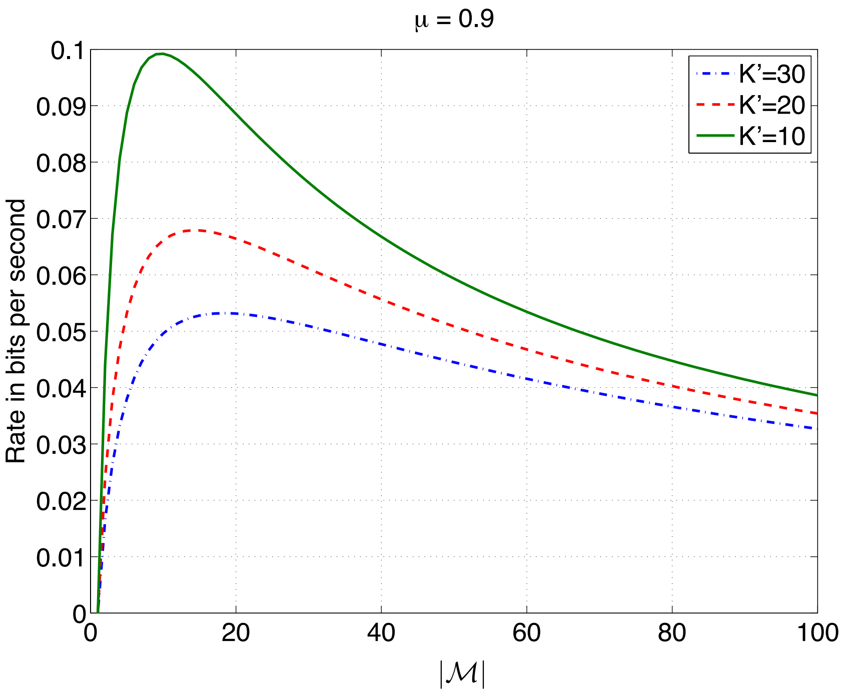


Fig. 6. Plot of the variation of rate as a function of the support set for M_i, which is $|\mathcal{M}|$. For a fixed parameter μ, the effective rate of transfer varies in the order $O(\frac{\log|\mathcal{M}|+K'}{|\mathcal{M}|+K'})$.

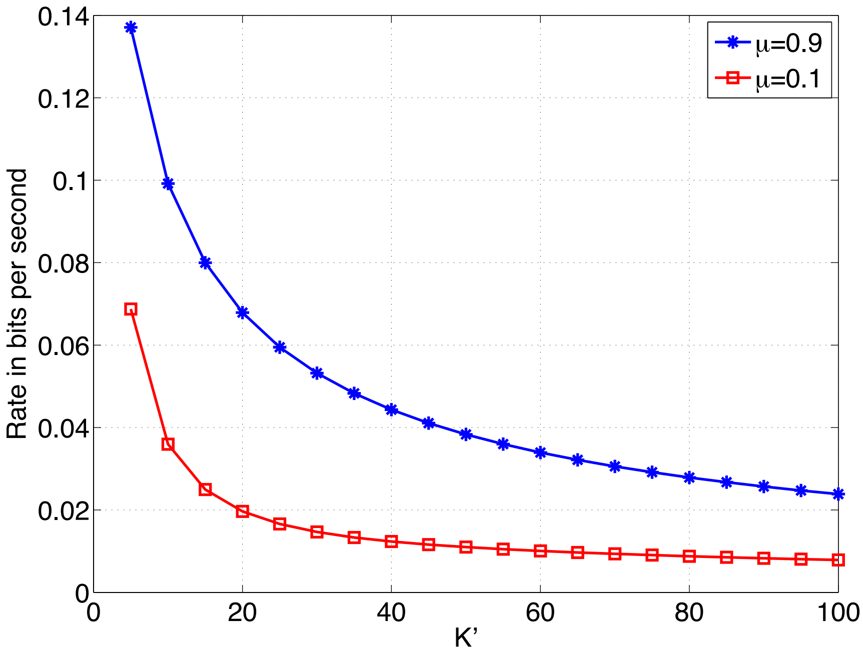


Fig. 7. Maximum rate at which HU can communicate to LU as a function of the support set of the noise, K'. This plot characterizes the 'effective rate of transfer' ($\tilde{R}_{trunc}$) vs the support of noise added by Pump to the acknowledgements sent to LU (K').

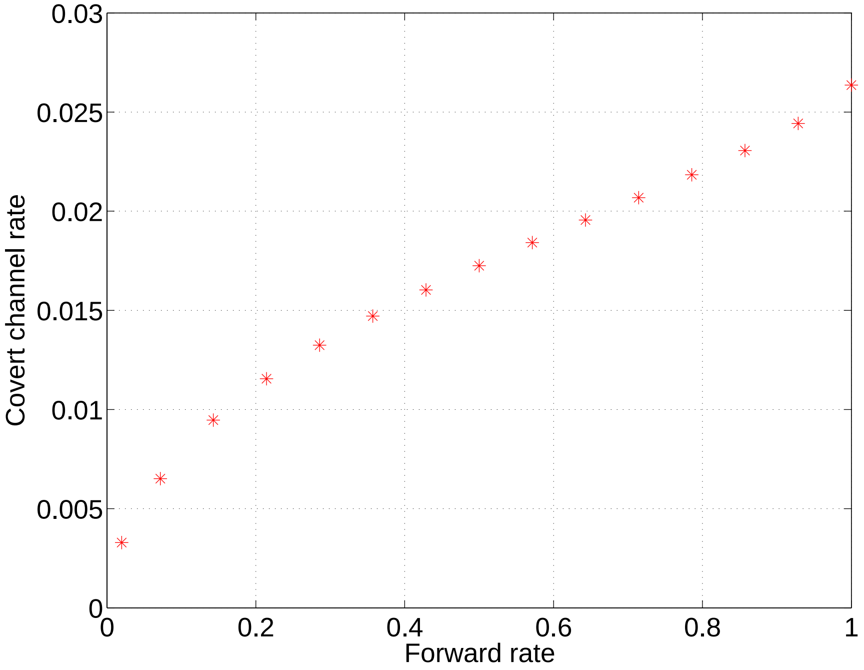


Fig. 8. The Forward rate is the rate at which the packets are transmitted from LU to HU. The covert communication rate is the rate from HU to LU. The support set of the noise was fixed at 100 for this simulation.

VII. Upper bound: HU to LU transmission rate

Let R be the rate of transmission (2) from HU to LU. The receiver (LU), determines and observes the sequences X^n and Z^n respectively, and uses only these sequences to estimate W. Let $P_e^{(n)}$ be the probability of making an error $\mathbb{P}\left(W \neq \hat{W}_n\right)$, where, $\hat{W}_n = f(X^n, Z^n)$ is the estimate of W at the receiver. We say that rate R is achievable if $\mathbb{P}\left(W \neq \hat{W}_n \to 0\right)$.

Theorem 7.1: If R is achievable, then $R \leq \mathcal{I}(\mathcal{Y}, \mathcal{B} \to \mathcal{Z} | B_0)$

Proof: Since R is achievable, Fano's inequality implies

$$H(W|X^n, Z^n) \leq 1 + P_e^{(n)} nR. \tag{28}$$

Hence, if $P_e^{(n)} \to 0$ as $n \to \infty$, then $\frac{1}{n}H(W|X^n, Z^n) \to 0$. Thus, if the error in decoding is to go to zero, then the uncertainty in W at the receiver should certainly go to zero.

$$\begin{aligned}
nR &= H(W) = H(W|B_0) && (29)\\
&= H(W|X^n, Z^n, B_0) + I(W; X^n, Z^n|B_0)\\
&= H(W|X^n, Z^n, B_0) + I(W; Z^n|B_0)\\
&\quad + I(W; X^n|Z^n, B_0)\\
&= H(W|X^n, Z^n, B_0) + I(W; Z^n|B_0) && (30)\\
&\leq H(W|X^n, Z^n) + I(W; Z^n|B_0) && (31)\\
&\leq 1 + P_e^{(n)} nR + I(W; Z^n|B_0) && (32)
\end{aligned}$$

where (29) follows because the message to be transmitted is assumed independent of the initial state of the buffer, (30) follows because X_n is a deterministic function of Z^n, (31) follows because conditioning reduces entropy, and finally, (32) follows from (28). Using Bayes' rule

$$I(W; Z^n|B_0) = \mathbb{E}\left[\log \frac{P_{W|Z^n,B_0}}{P_{W|B_0}}\right] = \mathbb{E}\left[\log \frac{P_{Z^n|W,B_0}}{P_{Z^n|B_0}}\right]. \tag{33}$$

$$\begin{aligned}
P_{Z^n|W,B_0} &= \prod_{i=1}^{n} P_{Z_i|Z^{i-1},W,B_0} && (34a)\\
&= \prod_{i=1}^{n} P_{Z_i|Z^{i-1},W,Y^i,B^i} && (34b)\\
&= \prod_{i=1}^{n} P_{Z_i|Z^{i-1},Y^i,B^i} && (34c)
\end{aligned}$$

where (34a) follows by chain rule, (34b) is because of Lemma 5.1, (34c) follows because our channel is non-anticipative (refer Figure 5), or equivalently, the Markov relation $W - (y^i, b^i, z^{i-1}) - z_i$ holds. With this, we have:

$$\begin{aligned}
I(W; Z^n|B_0) = &\sum_{i=1}^{n} \mathbb{E}\left[\log \frac{P_{Z_i|Z^{i-1},W}\left(Z_i|Z^{i-1}, W\right)}{P_{Z_i|Z^{i-1},B_0}\left(Z_i|Z^{i-1}, B_0\right)}\right]\\
&= \sum_{i=1}^{n} \mathbb{E}\left[\log \frac{P_{Z_i|Z^{i-1},Y^i,B^i}\left(Z_i|Z^{i-1}, Y^i, B^i\right)}{P_{Z_i|Z^{i-1},B_0}\left(Z_i|Z^{i-1}, B_0\right)}\right] && (35a)\\
&= \sum_{i=1}^{n} I(Y^i, B^i; Z_i|Z^{i-1}, B_0) = I(Y^n, B^n \to Z^n|B_0)
\end{aligned}$$

where (35a) follows from (33) and (34); Thus the information rate from the HU to LU is represented in terms of directed information flow over the system dynamics. From (32), (35)

$$nR \leq 1 + P_e^{(n)} nR + I(W; Z^n | B_0) \tag{36a}$$

$$= 1 + P_e^{(n)} nR + I(Y^n, B^n \rightarrow Z^n | B_0) \tag{36b}$$

Dividing by n and let $n \rightarrow \infty$, we conclude

$$R \leq \frac{1}{n} I(Y^n, B^n \rightarrow Z^n | B_0) = \mathcal{I}(\mathcal{Y}, \mathcal{B} \rightarrow \mathcal{Z} | B_0)$$

■

A. Significance of Directed Information Upper Bound:

We have so far shown that the directed information rate from (Y, b) to Z is an upper-bound on the rate of communication over this channel. The tightness of this bound is yet to be investigated and can only be verified by demonstrating an achievable scheme. The directed information expression over the channel can then be formulated as an equivalent dynamic programming problem and approximate solution can be found just like in the case of unifilar Finite State Channel (FSC) [34] and the feedback capacity problem [38]. The directed information bound is proven to be tight for similar examples of Trapdoor channel with feedback [34] and an Exponential Server Timing Channel (ESTC) [35]. We will discuss the example of ESTC where the directed information bound is tight.

Consider the communication system in Figure 9. An ESTC can be interpreted as a special case of such a system [36]. In an ESTC, where time is slotted finely enough to ensure that in a given time slot of duration Δ, there can be at most one arrival. X_k is the number of packets in the queue at time k, Z is the arrival process (chosen by the encoder), Y is the departure process, and $\mu(\cdot)$ is the update law of the queue, $X_k = X_{k-1} + Z_k - Y_{k-1}$. The memoryless channel is a Z-channel, $P(Y_k = 1 | X_k \neq 0) = \gamma\Delta$ and $P(Y_k = 1 | X_k = 0) = 0$, where γ is the mean service time of the ESTC. Similar to our problem, directed information is an upper bound to the communication rate in an ESTC, and also, a rate equal to directed information can be achieved over the channel when using Poisson inputs [35]. It is in essence because the dynamical system that is unknown to the encoder but does operate on previous outputs of the noisy channel, is in some sense optimally using that feedback, along with the encoder's input process. In the case of communication over the Pump, the encoder (HU) does have some information about the state (the encoder knows if the buffer B is empty or not), and hence, is in some sense even perhaps better off than the exponential server timing channel. We might therefore hope that the directed information bound reasonably approximates the actual rate, if not tightly, and our future efforts will be directed in developing achievable schemes.

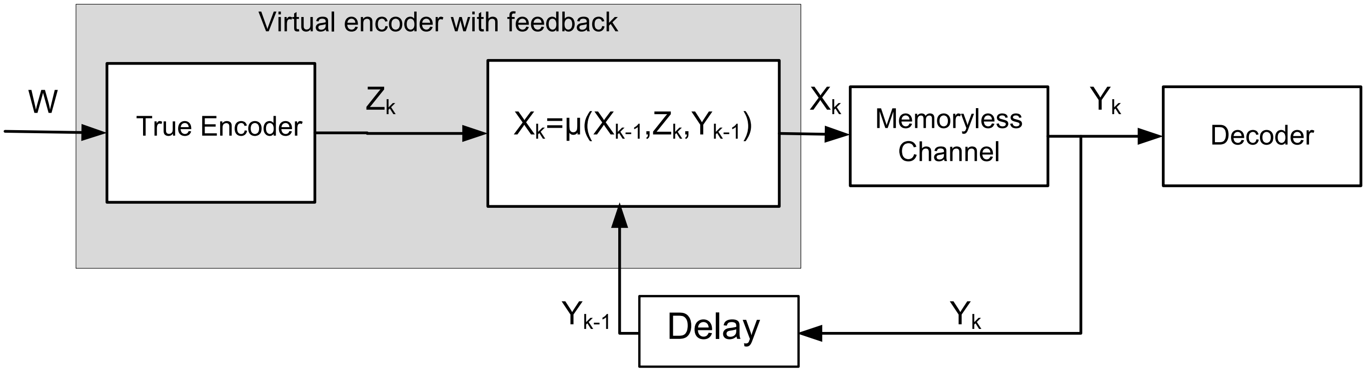


Fig. 9. Communication over a channel with dynamics

VIII. Conclusions, Discussion and Future Work

We have analyzed the capacity of the covert channel present when a Pump is used to isolate communication between two users (HU and LU) with different clearance levels, and the Pump buffer becomes full. Following a careful modeling of its working, we show that it is possible to communicate over this channel. The HU-LU communication is interpreted as communication over a channel with noisy feedback, and we provide theoretical bounds on the rate of communication. The upper bound is nonconstructive and is in terms of a directed information over the parameters of the system. However because of similarity of the problem at hand to the Trapdoor channel with feedback [34] and the Exponential Server Timing Channel (ESTC) [35], we will not surprised if in fact the directed information bound can be proven to be tight. In fact, proving the tightness of the upper bound is an interesting direction for future work.

Our lower bounds on the capacity of the channel between HU and LU are constructive and we present an achievability scheme which guarantees non-zero communication rate (infinite bits) if the noise added by the pump has a finite support. Even if the pump adds random noise with infinite support to the ACKs from Pump to LU still reliable communication is possible albeit at zero rate, i.e, finite number of bits can be transmitted. As depicted in Figure 7, the achievable communication rate is a function of support of noise added by pump. It is noteworthy that by increasing the support of noise, not only Pump can reduce the effective rate of communication between HU and LU, but also it will affect the the QoS of the system as it will slow down the legitimate communication rate between LU and HU. In short, depending on the QoS and security requirements of a system (the tolerable communication rate from HU to LU), Pump must vary its noise support.

IX. Acknowledgements

The authors would like to thank the reviewers for their useful comments and suggestions.

References

[1] B. Lampson, "A note on the confinement problem," in *Comm. ACM*, vol. 16, October 1973, pp. 613–615.

[2] M. H. Kang and I. S. Moskowitz, "A Pump for rapid, reliable, secure communication," in *Proc. ACM Conf. Computer & Comm. Security*, 1993, pp. 119–129.

[3] M. H. Kang, I. S. Moskowitz, and D. C. Lee, "A network pump," *IEEE Transactions on Software Engineering*, vol. 22, pp. 329–338, 1996.

[4] S. Chincheck, M. H. Kang, I. S. Moskowitz, , and J. Parsonese, "Systems and methods for providing increased computer security," in *US Patent 7,149,897 B2*, December 2006.

[5] D. E. Bell and L. LaPadula, "Secure computer systems," in *ESD-TR-73-278, Mitre Corp. V, 1, 2, & 3*, Nov. 1973 and April 1974.

[6] I. S. Moskowitz and M. H. Kang, "Covert channels - here to stay?" in *Compass'94: 9th Annual Conference on Computer Assurance. Gaithersburg, MD: National Institute of Standards and Technology*, 1994, pp. 235–244.

[7] I. S. Moskowitz and A. R. Miller, "Simple timing channels," in *Proc. IEEE Symp. on Research in Security & Privacy*, May 1994, pp. 56–64.

[8] S. Gorantla, S. Kadloor, T. Coleman, N. Kiyavash, I. Moskowitz, and M. Kang, "Directed Information and the NRL Network Pump," in *International Symposium on Information Theory and its Applications (ISITA)*, 2010.

[9] B. W. Lampson, "A note on the confinement problem," *Commun. ACM*, vol. 16, no. 10, pp. 613–615, 1973.

[10] J. K. Millen, "Covert channel capacity," in *IEEE Symposium on Security and Privacy*, 1987, pp. 60–66.

[11] J. C. Wray, "An analysis of covert timing channels," *Security and Privacy, IEEE Symposium on*, vol. 0, p. 2, 1991.

[12] W. M. Hu, "Reducing timing channels with fuzzy time," in *IEEE Symposium on Security and Privacy*. IEEE, 1991, pp. 8–20.

[13] ——, "Lattice scheduling and covert channels," in *IEEE Symposium on Security and Privacy*. IEEE, 1992, pp. 52–61.

[14] D. Brumley and D. Boneh, "Remote timing attacks are practical," *Computer Networks*, vol. 48, no. 5, pp. 701–716, 2005.

[15] P. Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," in *Advances in CryptologyCRYPTO96*. Springer, 1996, pp. 104–113.

[16] D. Chaum, "Blind signatures for untraceable payments," in *Advances in Cryptology: Proceedings of Crypto*, vol. 82, 1983, pp. 199–203.

[17] B. Köpf and G. Smith, "Vulnerability Bounds and Leakage Resilience of Blinded Cryptography under Timing Attacks," in *Computer Security Foundations Symposium (CSF), 2010 23rd*. IEEE, 2010, pp. 44–56.

[18] V. Anantharam and S. Verdu, "Bits Through Queues," *IEEE Trans. Inform. Theory*, vol. 42, pp. 4–18, Jan. 1996.

[19] J. Giles and B. Hajek, "An Information-Theoretic and Game-Theoretic Study of Timing Channels," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2455–2477, September 2002.

[20] A. Askarov, D. Zhang, and A. Myers, "Predictive black-box mitigation of timing channels," in *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 2010, pp. 297–307.

[21] S. Kadloor, X. Gong, N. Kiyavash, T. Tezcan, and N. Borisov, "A low-cost side channel traffic analysis attack in packet networks," in *IEEE ICC 2010 - Communication and Information System Security Symposium*, Cape Town, South Africa, 2010.

[22] J. Agat, "Transforming out timing leaks," in *Proceedings of the 27th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. ACM, 2000, pp. 40–53.

[23] A. Russo, J. Hughes, D. Naumann, and A. Sabelfeld, "Closing internal timing channels by transformation," *Advances in Computer Science-ASIAN 2006. Secure Software and Related Issues*, pp. 120–135, 2007.

[24] A. Di Pierro, C. Hankin, and H. Wiklicky, "Quantifying timing leaks and cost optimisation," *Information and Communications Security*, pp. 81–96, 2008.

[25] I. S. Moskowitz and A. R. Miller, "The channel capacity of a certain noisy timing channel," *IEEE Trans. Inform. Theory*, vol. 38, no. 4, pp. 1339–1344, July 1992.

[26] M. Kang, A. Moore, and I. Moskowitz, "Design and assurance strategy for the NRL pump," in *High-Assurance Systems Engineering Workshop, 1997., Proceedings*, Aug. 1997, pp. 64 –71.

[27] S. Gianvecchio and H. Wang, "Detecting covert timing channels: an entropy-based approach," in *Proceedings of the 14th ACM conference on Computer and communications security*. ACM, 2007, pp. 307–316.

[28] S. Cabuk, C. Brodley, and C. Shields, "IP covert channel detection," *ACM Transactions on Information and System Security (TISSEC)*, vol. 12, no. 4, pp. 1–29, 2009.

[29] H. Marko, "The bidirectional communication theory–a generalization of information theory," *IEEE Trans. Comm.*, pp. 1345–1351, Dec. 1973.

[30] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Information Theory Application (ISITA-90)*, 1990, pp. 303–305.

[31] G. Kramer, "Directed Information for Channels with Feedback," *ETH Series in Inform. Proc.,Konstanz: HartungGorre*, vol. 11, 1998.

[32] T. Cover and J. Thomas, *Elements of Information Theory*, 2006.

[33] S. Tatikonda and S. Mitter, "The Capacity of Channels With Feedback," *IEEE Trans. on Information Theory*, vol. 55, no. 1, pp. 323–349, 2009.

[34] H. Permuter, P. Cuff, B. V. Roy, and T. Weissman, "Capacity of Trapdoor Channel with Feedback," vol. 54, pp. 3150–65, July, 2008.

[35] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[36] S. K. Gorantla and T. P. Coleman, "Information-theoretic viewpoints on optimal causal coding-decoding problems," http://arxiv.org/abs/1102.0250, 2010.

[37] S. Gorantla and T. Coleman, "On reversible markov chains and maximization of directed information," *IEEE- Internation Symposium on Information Theory*, 2010.

[38] J. Chen and T. Berger, "The capacity of finite-state Markov Channels with feedback," in *IEEE Trans. on Info. Theory*, vol. 51, 2005, pp. 780–789.

Appendix A

Proof of Lemma 6.1

The probability of event E_1^n occurring is given by

$$\begin{aligned}
\mathbb{P}\left(E_1^n\right) &= \mathbb{P}\left(\cap_{i=1}^n E_{1,i} \cap E_0\right) \\
&= \left(\prod_{i=1}^n \mathbb{P}\left(E_{1,i} \middle| \cap_{k=1}^{i-1} E_{1,k}\right) \mathbb{P}\left(E_0\right)\right) && (37) \\
&= \left(\prod_{i=1}^n \mathbb{P}\left(M_i + N_i \leq K\right)\right) \mathbb{P}\left(E_0\right) && (38) \\
&\geq \left(\prod_{i=1}^n \mathbb{P}\left(N_i \leq K'\right)\right) \mathbb{P}\left(E_0\right) && (39) \\
&= \prod_{i=1}^n (1-(1-\mu)^{K'}) \mathbb{P}\left(E_0\right) && (40) \\
&= (1-(1-\mu)^{K'})^n (1-(1-\mu)^{K\bar{B}}) && (41) \\
&\to 1, \text{ as } n \to \infty && (42)
\end{aligned}$$

where (37) follows from the chain rule of probability, (39) holds true because by conditioning that all the receiver times $D_k \leq (\bar{B}+k)(K+1)$ till (i-1) transmissions, and thus does not interfere with the communication in the i-th block. In other words, $D_{i-1} \leq (\bar{B}+i)(K) < A_i < D_i$, and the receiver times $\{D_k\}_{k=1}^{i-1}$ does not interfere with the transmission in the i-th block. Thus, the communication for the i-th transmission is independent of the past. Hence $E_{1,i}$ conditioned on $E_{1,k}, k = 1, 2, \cdots, (i-1)$ having occurred is equivalent to $M_i + N_i \leq K$. (39) follows because $N_i \leq K' \implies M_i + N_i \leq K$, and so $\mathbb{P}(M_i + N_i \leq K) \geq \mathbb{P}(N_i \leq K')$. (40) follows from the properties of geometric noise (15). (42) holds if $K' > \frac{\log(n)}{|\log(1-\mu)|}$, i.e., K' grows at a rate greater than $\frac{\log n}{|\log(1-\mu)|}$.

Similarly, we can compute the conditional joint distribution of noise N_i conditioned upon event E_2^n using the conditional independence of communication between blocks when $N_i \leq K'$.

$$\begin{aligned}
&\mathbb{P}(N_1 = a_1, N_2 = a_2, \cdots, N_n = a_n | E_2^n) \\
&= \mathbb{P}(N_1 = a_1, N_2 = a_2, \cdots, N_n = a_n | E_2^n \cap E_1^n) && (43)\\
&= \prod_{i=1}^{n} \mathbb{P}(N_i = a_i | N_1 = a_1, \cdots, N_{i-1} = a_{i-1}, E_2^n, E_1^n) && (44)\\
&= \prod_{i=1}^{n} \mathbb{P}(N_i = a_i | E_2^n, E_1^n) && (45)\\
&= \prod_{i=1}^{n} \frac{\mathbb{P}(N_i = a_i, E_2^n, E_1^n)}{\mathbb{P}(E_2^n \cap E_1^n)} && (46)\\
&= \prod_{i=1}^{n} \frac{(1-\mu)^{a_i-1}\mu}{(1-(1-\mu)^{K'})} && (47)\\
&= \begin{cases} \frac{\prod_{i=1}^{n}(1-\mu)^{a_i-1}\mu}{\left((1-(1-\mu)^{K'})\right)^n}, & a_i \leq K', \forall\, i = 1, 2, \cdots, n \\ 0, & \text{otherwise.} \end{cases} && (48)
\end{aligned}$$

where (43) follows because $E_2^n \subset E_1^n$ by definition of E_2^n and E_1^n. (44) follows from chain law of probability, (45) follows from the independence between blocks conditioned upon E_2, (46) follows from conditional law of probability, (47) follows from the properties of geometric distribution. Hence, the noise is equivalent to a truncated geometric noise of rate μ and truncated at K' when conditioned upon event E_2^n.

APPENDIX B
PROOF OF LEMMA 6.2

The proof of (21a) and (21b) follows from standard coding theorem [32, Chapter 8]. For (21c),

$$
\begin{aligned}
&\mathbb{P}\left(\hat{W} \neq W | E_2^n\right) \\
&= \frac{\mathbb{P}\left(\hat{W} \neq W, E_2^n\right)}{\mathbb{P}\left(E_2^n\right)} && \text{(49a)}\\
&\leq \frac{\mathbb{P}\left(\hat{W} \neq W\right)}{\mathbb{P}\left(E_2^n\right)} && \text{(49b)}\\
&\leq \frac{\epsilon_2(n)}{\mathbb{P}\left(E_2^n\right)} && \text{(49c)}\\
&= \frac{\epsilon_2(n)}{(1-(1-\mu)^{K'})^n(1-(1-\mu)^{KB})} \triangleq \epsilon_3(n) && \text{(49d)}\\
&\rightarrow 0, \text{ as } n \rightarrow \infty, \text{ if } K' = \mathrm{O}\big(logn\big) && \text{(49e)}
\end{aligned}
$$

where (49a) follows from conditional law of probability, (49b) is because probability of an intersection of events is less than a single event, (49c) follows from (21b), (49d) follows from properties of geometric distribution and (49e) holds when $K' > log_{1/(1-\mu)}(n)$.

APPENDIX C
PROOF OF THEOREM 6.4

(25a) follows from (21a) of Lemma 6.2. (25b) can be proved as follows:

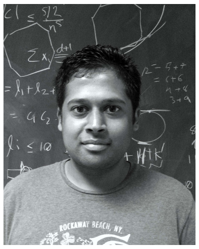
$$
\begin{aligned}
&\mathbb{P}\left(\hat{W} \neq W\right) \\
&= \mathbb{P}\left(\hat{W} \neq W | E_2\right)\mathbb{P}\left(E_2\right) + \mathbb{P}\left(\hat{W} \neq W | E_2^c\right)\mathbb{P}\left(E_2^c\right) && \text{(50a)}\\
&\leq \mathbb{P}\left(\hat{W} \neq W | E_2^n\right)\mathbb{P}\left(E_2^n\right) + \mathbb{P}\left((E_2^n)^c\right) && \text{(50b)}\\
&\leq \epsilon_3(n)\mathbb{P}\left(E_2^n\right) + \mathbb{P}\left((E_2^n)^c\right) && \text{(50c)}\\
&= \epsilon_3(n)(1-(1-\mu)^{K'})^n \\
&\quad + \left(1-(1-(1-\mu)^{K'})^n\right) \triangleq \epsilon_4(n) && \text{(50d)}\\
&\rightarrow 0, \text{ as } n \rightarrow \infty, && \text{(50e)}
\end{aligned}
$$

where (50a) follows from the additive law of probability, (50b) follows because probability of any event is less than 1, (50c) follows from (21c) in Lemma 6.2, (50d) follows from the properties of the geometric distribution where $K' = K - |\mathcal{M}|$, (50e) holds true if $K \geq \frac{\log(n)}{|\log(1-\mu)|} + |\mathcal{M}|$.

Siva K. Gorantla Siva K. Gorantla received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 2007 and the M.S. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 2009, where he is currently pursuing the Ph.D degree.

He was an intern at the Research Laboratory for Electronics at MIT, Adchemy Inc., and at University of New Mexico. His research interests include stochastic control, information theory, statistical learning and prediction.

Mr.Gorantla has been awarded the James Henderson fellowship at UIUC and S Subramanian Award (Institute Merit Prize) from IIT Madras in the years 2008 and 2004.



Sachin Kadloor Sachin Kadloor (S'07) graduated from Indian Institute of Technology Madras with a bachelors degree in electrical engineering in 2007. He then moved to University of Toronto where he graduated with a masters degree in 2009. His research then dealt with power allocation in selection based cooperative cellular networks. Since September 2009, he has been working towards his Ph.D. at the University of Illinois, Urbana-Champaign. His current research interests are information theory pertaining to timing channels and issues in network security.

Negar Kiyavash Negar Kiyavash joined the Department in January 2009. She received a B.S. degree in Electrical and Computer Engineering from the Sharif University of Technology, Tehran, in 1999, M.S. and Ph.D. degrees, also in Electrical and Computer Engineering, both from University of Illinois at Urbana-Champaign in 2003 and 2006, respectively.

From 2006 through 2008, Dr. Kiyavash was a Research Faculty at Department of Computer Science and a Research Scientist at Information Trust Institute both at the University of Illinois at Urbana-Champaign. Her research interests are in information theory and statistical signal processing with applications to computer, communication and multimedia security .

Todd P. Coleman Todd P. Coleman received the B.S. degrees in electrical engineering (summa cum laude), as well as computer engineering (summa cum laude) from the University of Michigan, Ann Arbor, in 2000, where he was awarded the College of Engineering's Hugh Rumler Senior Class Prize. He received the the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002, and 2005, where was a recipient of the National Science Foundation Graduate Research Fellowship and the MIT EECS Department's Morris J. Levin Award for Best Masterworks Oral Thesis Presentation. During the 2005-2006 academic year, he was a postdoctoral scholar at MITs Department of Brain and Cognitive Sciences and Massachusetts General Hospitals Neuroscience Statistics Research Laboratory in computational neuroscience. From July 2006-2011, he was an Assistant Professor in the ECE Department and Neuroscience Program at the University of Illinois. Since July 2011, he has been an Associate Professor in Bioengineering at the University of California, San Diego. His research interests include information theory, control theory, and neuroscience with a common theme being the interaction between multiple decision-makers. In Fall 2008, he was a co-recipient of the University of Illinois College of Engineerings Grainger Award in Emerging Technologies. While at U. Illinois, he also served as a co-Principal Investigator for an NSF IGERT interdisciplinary training grant for graduate students, titled ”Neuro-engineering: A Unified Educational Program for Systems Engineering and Neuroscience.

Ira S. Moskowitz Dr. Moskowitz earned his PhD in Mathematics from Stony Brook University. Along with Dr. Kang, Dr. Moskowitz is the co-inventor of the patented NRL Network Pump. He has been at NRL for 22 years. He specializes in the area of covert channel analysis via Information Theory. He has also done research in steganography and watermarking.

Myong H. Kang Dr. Myong H. Kang holds a PhD in Electrical Engineering from Purdue. His speciality is software security. His current research interests include secure mobile devices, secure cloud computing, computer network defenses , and he is presently the head of the computer security section of NRL's Center for High Assurance Computer Systems.